"Express Mail" mailing label number EV 320858119 US

Date of Deposit: 11/19/03

Our Case No. 09800240-0055

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
APPLICATION FOR UNITED STATES LETTERS PATENT

INVENTORS: Stephen A. Boppart
Daniel L. Marks

TITLE: NONLINEAR INTERFEROMETRIC VIBRATIONAL IMAGING

REPRESENTATIVE: Paul E. Rauch, Ph.D., Reg. No. 38,591
William J. Keyes, Ph.D., Reg. No. 54,218
SONNENSCHEIN NATH & ROSENTHAL LLP
P.O. Box #061080
Wacker Drive Station
Sears Tower
Chicago, Illinois 60606-1080
312-876-8000

NONLINEAR INTERFEROMETRIC VIBRATIONAL IMAGING

CROSS REFERENCE TO RELATED APPLICATIONS

The present application claims priority to U.S. Provisional application 60/442,300, filed January 24, 2003, which is hereby incorporated by reference.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

The subject matter of this application may in part have been funded by the NASA/National Cancer Institute (Contract no. NAS2-02057). The government may have certain rights in this invention.

BACKGROUND

The ability to measure the three-dimensional structure of biological tissues is important, but common methods used to leave out basic information about the molecular composition and metabolic behavior of the tissue imaged. This molecular composition and metabolic behavior information may yield valuable scientific data on the behavior of biological systems, and would be of great clinical diagnostic value for finding diseases such as cancer. Much of the focus of biological and medical imaging today is to gain information about composition.

Functional magnetic resonance imaging (FMRI) utilizes contrast agents specific to particular molecular species or metabolic processes to provide specific information about the three-dimensional (3-D) location of these species or processes. This method is quite versatile, and poses relatively little risk to the patient, but is limited in practice to a resolution over 100 microns. Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT) utilizes a radiolabeled metabolic molecule to provide 3-D measurements of utilization of these molecules by sensing their gamma-ray emissions. Unfortunately, these methods have been cost prohibitive as diagnostic tools, and expose the patient to ionizing radiation. Fluorescence microscopy labels relevant biological structures and processes

with a fluorescent compound which can be measured by conventional microscopy or confocal microscopy. The two-photon variant uses ultrafast pulses so that emission and excitation frequencies can be clearly separated, and also utilizes the additional selectivity that the intensity-sensitive nature of two-photon excitation provides. While some biological structures produce natural fluorescence, in general externally introduced fluorescence markers must be used, which can interfere with biological processes and are often toxic.

Molecules frequently have molecular resonance frequencies that are due to the electromagnetic attractions of atoms in the molecule. These frequencies are those of molecular vibrations, molecular rotational motions, the excitation of electrons to higher energy states, and occasionally finer structures such as hyperfine interactions and optical-magnetic properties. These properties are present without the introduction of any external contrast molecule. These frequencies are usually in the mid-infrared, corresponding to photons of 1.5-50 microns of wavelength. Unfortunately, they cannot be directly excited by electromagnetic radiation of the same frequency because when they are in tissue, the surrounding water absorbs almost all of these frequencies. The range of wavelengths that the tissue is relatively transparent to is 0.6-1.5 microns. Therefore multiphoton nonlinear processes need to be employed to probe these resonances. The photons to stimulate and record the processes are typically in a region where the tissue is not absorbing, so that they can reach the tissue feature and be measured from the feature.

Raman spectroscopy, first discovered in 1928, uses molecular resonance features of frequency $\Delta\omega$ to split a photon of frequency ω into another photon of frequency $\omega - \Delta\omega$ and a resonance excitation of frequency $\Delta\omega$. The presence of photons at frequency $\omega - \Delta\omega$ identifies the concentration of the resonance feature. This process is in practice very weak and requires large amounts of power to produce any detectable amounts of photons. This weakness is due to the fact that the probability of a Raman excitation process to occur is proportional to the number of photons at frequency $\omega - \Delta\omega$ already present, of which there are typically few or none.

Since photons that would be emitted by Raman excitation at frequency $\omega - \Delta\omega$ are indistinguishable from the incoming radiation that stimulates them, this is not a viable technique for achieving molecular sensitivity.

Coherent Anti-Stokes Raman Scattering (CARS) is another nonlinear spectroscopy technique that unlike conventional Raman spectroscopy, allows all of the photons necessary to stimulate the process to be introduced into the tissue by the illuminating source. This enables the probability of a CARS interaction to be increased to a (theoretically arbitrarily) high level so that a sufficient number of photons can be produced as to enable detection within a reasonable time period. It is essentially two stimulated Raman scattering processes in parallel. Two photons, a "pump" of frequency ω_1 and a "Stokes" of frequency ω_2 illuminate the tissue. They must be separated in frequency by $\omega_1 - \omega_2 = \Delta\omega$, which is the frequency of the molecular resonance. When molecules of the target molecular species are present, the resonance will be excited, and the pump photon will be converted to the same frequency as the Stokes photon. This is the first stimulated Raman scattering process. Another photon may arrive at frequency ω_3 that will stimulate the emission of the excitation from the resonance, so that the energy of the photon of frequency ω_3 and the excitation are converted to a new photon of frequency $\omega_4 = \omega_3 + \Delta\omega$, called the "anti-Stokes" photon. The presence of this photon of frequency ω_4 indicates that a CARS process has taken place and indeed a molecule with the resonance feature is present. Often the "pump" beam is used as the photons of frequency ω_3, so that $\omega_3 = \omega_1$ and $\omega_4 = 2\omega_1 - \omega_2$. Since the photon of ω_4 is not the same frequency as one of the illuminating photons, and is typically within the transparency range of the tissue, it is easily discriminated from the incoming radiation. Figure 1 A shows an energy-level diagram for CARS, and Figure 1 B shows an energy-level diagram for Coherent Stokes Raman Scattering.

CARS microscopy uses the CARS process to look for the presence of a molecular species, but does not require any foreign substances to be

introduced into the tissue. It scans the illumination point-by-point through the tissue and measures the number of generated anti-Stokes photons. When a three-dimensional mesh of points has been scanned, a complete three-dimensional picture of molecules of that resonance can be shown. Since CARS is a nonlinear process (and therefore is intensity sensitive), efficient conversion only occurs at the focus of the illumination, which can be made very tight (typically a half micron in both the axial and lateral directions). Therefore the resolution can be made many orders of magnitude better than MRI, which is the probably the largest competition for clinical use for similar purposes. Unfortunately, the penetration is usually rather low (limited to about 500 microns). A further shortcoming is that CARS microscopy measures the total number of anti-Stokes photons, or power, from the sample. However, the optical field contains temporal structure in the phase that is averaged out by power detection because photodetector response time is orders of magnitude slower than the oscillations of the optical field. The time scale on which the optical pulse varies (which is typically picoseconds or femtosecond time scales) is far too fast for photon detection equipment or electronics to detect (the fastest of which may detect 25 ps time scales).

Optical coherence tomography (OCT) is an emerging high-resolution medical and biological imaging technology [15-21]. OCT is analogous to ultrasound B-mode imaging except reflections of low-coherence light are detected rather than sound. OCT detects changes in the backscattered amplitude and phase of light.

Cross-sectional OCT imaging is performed by measuring the backscattered intensity of light from structures in tissue. This imaging technique is attractive for medical imaging because it permits the imaging of tissue microstructure *in situ*, yielding micron-scale imaging resolution without the need for excision and histological processing. Because OCT performs imaging using light, it has a one- to two-order-of-magnitude higher spatial resolution than ultrasound and does not require contact with tissue.

OCT was originally developed and demonstrated in ophthalmology for high-resolution tomographic imaging of the retina and anterior eye [22-24].

Because the eye is transparent and is easily optically accessible, it is well-suited for diagnostic OCT imaging. OCT is promising for the diagnosis of retinal disease because it can provide images of retinal pathology with 10 μm resolution, almost one order-of-magnitude higher than previously possible using ultrasound. Clinical studies have been performed to assess the application of OCT for a number of macular diseases [23,24]. OCT is especially promising for the diagnosis and monitoring of glaucoma and macular edema associated with diabetic retinopathy because it permits the quantitative measurement of changes in the retinal or retinal nerve fiber layer thickness. Because morphological changes often occur before the onset of physical symptoms, OCT can provide a powerful approach for the early detection of these diseases.

Recently, OCT has been applied for imaging a wide range of nontransparent tissues [16,17,25-27]. In tissues other than the eye, the imaging depth is limited by optical attenuation due to scattering and absorption. A "biological window" exists in tissue where absorption of near-infrared wavelengths is at a minimum and light can penetrate deep into highly-scattering tissue (Figure 15) [28]. Because optical scattering decreases with increasing wavelength, OCT in nontransparent tissues has routinely used 1.3 μm wavelength light for imaging. In most tissues, imaging depths of 2-3 mm can be achieved using a system detection sensitivity of 110 dB (1 part in 10^{11}). OCT has been applied to image arterial pathology *in vitro* and has been shown to differentiate plaque morphology with superior resolution to ultrasound [17,29].

Imaging studies have also been performed to investigate applications in gastroenterology, urology, and neurosurgery [30-32]. High resolution OCT using short coherence length, short-pulse light sources, has also been demonstrated and axial resolutions of less than 5 μm have been achieved [33,34]. High-speed OCT at image acquisition rates of 4 to 8 frames per second for 500 to 250 square pixel images has been achieved [35]. OCT has been extended to perform Doppler imaging of blood flow and birefringence imaging to investigate laser intervention [36-38]. Different imaging delivery

systems including transverse imaging catheters and endoscopes, and forward imaging devices have been developed to enable internal body OCT imaging [39,40]. Most recently, OCT has been combined with catheter-endoscope-based delivery to perform *in vivo* imaging in animal models and human patients [41-44].

Apart from medical applications, OCT has been demonstrated as an emerging investigational tool for cell and developmental biology. OCT has imaged the development of numerous animal models including *Rana pipiens* and *Xenopus laevis* (Leopard and African frog), and *Brachydanio rerio* (zebrafish) [45-46]. High-speed OCT imaging has permitted the morphological and functional imaging of the developing *Xenopus* cardiovascular system, including changes in heart function following pharmacological interventions [47]. High-resolution imaging has permitted the real-time tracking of cell dynamics in living specimens including mesenchymal cell mitosis and neural crest cell migration [48]. OCT is advantageous in microscopy applications because repeated non-invasive imaging of the morphological and functional changes in genetically modified animals can be performed overtime without having to histologically process multiple specimens. The high-resolution, cellular-imaging capabilities suggest that OCT can be used to diagnose and monitor early neoplastic changes in humans.

The ability of OCT to perform optical biopsies, the *in situ* imaging of tissue microstructure at near-histological resolution, has been used to image morphological differences between normal and neoplastic tissue. OCT images of *in vitro* neoplasms of the female reproductive tract [49], the gastrointestinal tract [50], and the brain [51] have been investigated. Optical differences between normal and neoplastic tissue were evident, but primarily for late-stage changes. Still, situations exists were no inherent optical contrast exists between normal and pathologic tissue, such as in early-stage, pre-malignant tumors or in tumors which remain optically similar to normal tissue.

In the past, OCT has found numerous medical and biological applications. However, the imaging technique has relied largely on the inherent optical properties of the tissue to provide contrast and differentiate normal from pathological tissue. Phospholipid-coated perfluorobutane microbubbles (ImaRx Pharmaceutical, Tucson, AZ) have been used as a contrast agent for OCT; although they produce a strong OCT signal, blood and tissue also produce a fairly strong OCT signal, and the effects of this contrast agent *in vivo* on the visualization of blood vessels are subtle.

SUMMARY OF THE INVENTION

In a first embodiment, the invention provides a method of examining a sample, which includes: exposing a reference to a first set of electromagnetic radiation, to form a second set of electromagnetic radiation scattered from the reference; exposing a sample to a third set of electromagnetic radiation to form a fourth set of electromagnetic radiation scattered from the sample; and interfering the second set of electromagnetic radiation and the fourth set of electromagnetic radiation. In this embodiment, the first set and the third set of electromagnetic radiation are generated from a source. Moreover, at least a portion of the second set of electromagnetic radiation is of a frequency different from that of the first set of electromagnetic radiation, and at least a portion of the fourth set of electromagnetic radiation is of a frequency different from that of the third set of electromagnetic radiation.

In a second embodiment, the invention provides a method of forming an image of a sample, which includes: exposing a reference to a first set of electromagnetic radiation, to form a second set of electromagnetic radiation scattered from the reference; exposing a sample to a third set of electromagnetic radiation to form a fourth set of electromagnetic radiation scattered from the sample; forming a digital data set corresponding to the sample; and converting the data set into an image. In this embodiment, the data set is formed by interfering the second set of electromagnetic radiation and the fourth set of electromagnetic radiation. Also, the first set and the third set of electromagnetic radiation are generated from a source. Moreover, at

least a portion of the second set of electromagnetic radiation is of a frequency different from that of the first set of electromagnetic radiation, and at least a portion of the fourth set of electromagnetic radiation is of a frequency different from that of the third set of electromagnetic radiation.

In a third embodiment, the invention provides a device for examining of a sample having an oscillator, a reference generator that is optically coupled to the oscillator, a microscope that is optically coupled to the oscillator, a demodulator that is optically coupled to the reference generator and the microscope, and a recorder that is coupled to the demodulator.

In a fourth embodiment, the invention provides a method of examining a sample that includes exposing a sample to a first set of electromagnetic radiation to form a second set of electromagnetic radiation scattered from the sample, and interfering the second set of electromagnetic radiation with a third set of electromagnetic radiation. In this embodiment, the third set of electromagnetic radiation is phase-coherent with the first set of electromagnetic radiation, at least a first portion of the second set of electromagnetic radiation is of a frequency different from that of the first set of electromagnetic radiation, and at least a portion of the third set of electromagnetic radiation is of the same frequency as the first portion of the second set of electromagnetic radiation.

In a fifth embodiment, the invention provides a method of forming an image of a sample that includes exposing a sample to a first set of electromagnetic radiation to form a second set of electromagnetic radiation scattered from the sample, forming a digital data set corresponding to the sample, and converting the data set into an image. In this embodiment, the forming of the image includes interfering the second set of electromagnetic radiation and a third set of electromagnetic radiation. Moreover, the third set of electromagnetic radiation is phase-coherent with the first set of electromagnetic radiation. In addition, at least a first portion of the second set of electromagnetic radiation is of a frequency different from that of the first set of electromagnetic radiation, and at least a portion of the third set of

electromagnetic radiation is of the same frequency as the first portion of the second set of electromagnetic radiation.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Coherent Anti-Stokes Raman Scattering and Coherent Stokes Raman Scattering energy-level diagrams.

Figure 2. Basic block diagram of NIVI

Figure 3. Example laser configurations are shown that produce two pulses of frequencies ω_1 and ω_2 overlapped. We assume that the delays have been set correctly to overlap them.

Figure 4. Methods of shaping a broadband pulse into a pulse with beat frequency $\Delta\omega$.

Figure 5. Reference Signal Generator Implementation

Figure 6. Configurations for full field CARS.

Figure 7. Translated serial-point scanning configurations

Figure 8. Beam-steered serial-port scanning configurations

Figure 9. Temporal-ranging based NIVI

Figure 10. Full field cross-correlator demodulator

Figure 11. Temporal cross-correlator for a serial-point scanning microscope.

Figure 12. Shown are two configurations that utilize linear photodetector arrays to measure multiple samples of the cross-correlation simultaneously.

Figure 13. Diagram of an implementation of the recorder.

Figure 14. Electrical cross-correlation signal quadrature amplitude demodulator.

Figure 15. "Biological window" in tissue where absorption of near-infrared wavelengths is at a minimum and light can penetrate deep into highly-scattering tissue.

DETAILED DESCRIPTION

Nonlinear interferometric vibrational imaging (NIVI) is a method used to measure the three-dimensional distribution of molecular species in various samples (biological or otherwise). Its basic operation is to stimulate the excitation of molecular bonds with particular resonance frequencies, and then use these excitations to produce radiation distinct from the excitation that can be measured. The physical process of excitation and stimulation of radiation is called Coherent Anti-Stokes Raman Scattering (CARS). Unlike previous methods that use CARS in microscopy to probe for the presence of molecular species, NIVI utilizes a heterodyne approach where a reference signal is separately generated and interferomerically compared to the signal received from the sample, allowing the signal to be formed into an image in the same way as OCT. In this way, additional information can be inferred from the emitted radiation such as the distance to the sample and phase information that yields additional structure of the molecular bonds. It also has other advantages in sensitivity and the ability to screen out background radiation that is not produced by the sample. It also can allow more flexibility in the choice of laser illumination source, because the coherent detection process does not rely on photon frequency alone to discriminate emitted radiation.

The term "scattered photon" means photons scattered by a sample, which include linearly scattered photons and non-linearly scattered photons, such as CARS and CSRS photons.

The term “image” means data produced by receipt of electromagnetic radiation, which may or may not be formed into a picture viewable by the human eye. This includes images produced directly onto a medium such as film or video.

The phrase “frequency range of infra-red to ultraviolet” means electromagnetic radiation having a frequency of 10^{12} to 10^{17} Hz, which excludes radio waves, microwaves, X-rays and gamma rays. The term “light” means visible light.

The method of NIVI was developed at first because CARS microscopy based on non-interferometric detection is incompatible with the scanning

modes of OCT. When the method of CARS microscopy is employed, information about the echo time of the optical signal is lost because the photodetector is not nearly fast enough to respond to the relatively instantaneous return of the return pulse. In OCT a reference wave is used so that the relative time delay between the reference wave and the returned echo wave can be determined. Nonlinear methods (such as conventional Raman spectroscopy) are unsuited to integration with OCT because the emitted signal is incoherent with respect to the excitation. However, CARS processes preserve coherence, and therefore enable interferometric methods to be used. By using a nonlinearly generated reference wave, the time of arrival of a CARS signal returning from the tissue can be interferometrically compared to the arrival time of the reference signal. In this way, the methods of OCT and CARS imaging can be integrated. In addition, in a manner analogous to how OCT can measure the dispersion of scatterers in the medium, NIVI can measure the dispersion of the response of molecules to the excitation radiation. This is because NIVI measures the relative phase between the reference signal and the sample signal. This dispersion should contain information about the resonance structure of the molecule over and above what can be measured using CARS microscopy.

The Nonlinear Interferometric Vibrational Imaging Method

The purpose of NIVI is to measure the temporal field radiated by samples that are stimulated using the CARS technique. When the field is scanned through the sample, an image can be formed of the molecular contents of the sample. As illustrated in Figure 2, a preferred embodiment NIVI **200** has the following components:

Oscillator. This oscillator **201** produces an optical field that can excite the resonance mode of the target molecule through a nonlinear technique (usually stimulated Raman scattering), and also the photon to stimulate the photon that is measured (also usually through stimulated Raman scattering). The combination of these processes is called CARS.

Reference Generator. The reference signal generator **203**, which can sometimes be incorporated into the oscillator, converts part of the oscillator signal to a reference signal that can be used in the demodulator section. It acts as a known signal that demodulates the unknown signal from the sample in the interferometer.

Microscope. The microscope **205** delivers the field produced by the oscillator to the sample, and collects the field emitted by the sample. The excitation field is usually delivered by a microscope objective, where the oscillator field is focused tightly or sparsely, depending on the scanning method. This focus is scanned through the tissue, and based on the signal received from each tissue volume an image can be formed. When the oscillator signal is delivered to the tissue containing a molecule with a compatible resonance, a nonlinear process such as CARS can occur and produce a new sample signal (called the "anti-Stokes" for CARS processes). This sample signal serves as an indicator of the presence of the molecular resonance, and also provides additional information about the molecule through the temporal structure of the sample signal.

Demodulator. The demodulator **207** combines the signal received from the sample with the reference signal. This is typically achieved by constructing an interferometric cross-correlator. The cross-correlation of the two signals is then measured by a single photodetector or array of photodetectors. The power received by these photodetectors allows the cross-correlation signal to be inferred, from which the temporal response signal from the sample can be also inferred. With knowledge of the physics of the molecule, the presence of and potentially properties of the molecule being tested can be inferred from its temporal response.

Recorder. The data recorder **209** records the data measured by the demodulator. This data can be digitally processed to produce an image that a human operator can interpret.

Each of these modules can be implemented in a variety of different ways that can be tailored to various data acquisition needs. In addition, while these units are the basic units of the invention, often the parts can be

consolidated to simplify implementation or reduce cost. While the basic block design could be construed as that of a standard interferometric microscope, nonlinear processes are occurring in the reference generator and "Microscope" sections that allow the resonance information of the sample to be pumped.

Each of these units will be detailed presently.

1. **Oscillator.**

The oscillator produces the electromagnetic field that stimulates the excitation of the resonance to the probed. It also provides the photon that stimulates the output photon that is detected as evidence of CARS or CSRS. There are many types of oscillators and fields that can produce CARS. Each pulse produced by the oscillator should be nearly identical so that it can excite consistent signals in the reference generator and sample. If the oscillator produces too variable of a signal, the signals from the reference generator and sample may change and produce signals that can be confused with noise sources. Variability in the oscillator output is a noise source in itself that adds uncertainty to what the expected demodulated signal should be.

The conventional way to produce CARS is to send in two overlapped optical pulses, one of which at frequency ω_1, the pump, and the other at ω_2, the Stokes pulse, where $\omega_1 - \omega_2 = \Delta\omega$, where $\Delta\omega$ is the resonance frequency of the molecules of interest. These pulses produce a beat frequency of $\Delta\omega$ that manifests itself in the magnitude of the optical field. In linear time-independent optics, systems are sensitive only to the frequencies of the optical pulses themselves, and not any beats they may produce together. However, with sufficient intensity the intensity envelope may itself stimulate the molecule. By choosing two pulses that produce beats of this frequency, we can stimulate the molecule with two wavelengths that the tissue is transparent to. Once the resonance is stimulated, another photon of frequency ω_1 (in CARS), or of frequency ω_2 (in CSRS) stimulates the emission of a fourth photon, which is of frequency $2\omega_1 - \omega_2$ for CARS, and $2\omega_2 - \omega_1$ for CSRS.

Systems that can be used to produce these two frequencies are shown below in Figure 3. A common configuration to produce pulses of these two wavelengths that are overlapped in time is to have a pulsed laser produce one of these wavelengths, split off of some of its energy, and use this energy to produce a second pulse of a lower or higher frequency. In one configuration **300**, a pump laser **301** pumps a dye laser **303**, for example a doubled Nd:YAG pump laser at 532 nm pumping a tunable dye laser. In another possible configuration **302**, a pump laser such as a Ti-sapphire oscillator pumps an optical parametric oscillator (a device that converts pulses to lower frequencies) **305**.

In yet another configuration **304**, the pump laser pumps a regenerative amplifier **307**, such as a Ti-sapphire regenerative amplifier. The regenerative amplifier then pumps an optical parametric amplifier (another frequency conversion device) **309**. Alternatively, as illustrated in configuration **306**, the pulses of each wavelength are generated by two separate pump lasers, and the time overlap is maintained by a circuit **311** that synchronizes the two sources. In another configuration **308**, the pump laser pumps a continuum light generator **313**, generating broadband light which is filtered by a filter for the two wavelengths with group velocity dispersion correction **315**.

While directly generating the two frequencies and superimposing them to produce beats is the most common way to stimulate CARS, this method has some disadvantages for the method of NIVI. In CARS and CSRS, there are two types of generated signals. Resonant signals depend on the presence of a molecule of a particular resonance frequency to be present to generate the CARS signal. Another component, nonresonant CARS, does not require a particular frequency to perform conversion. Nonresonant CARS depends on the peak intensity in the signal, while the resonant component can build up from many beat periods and so therefore can be spread out in time. Because of this, it is advantageous to spread the CARS signals in time to reduce the nonresonant component.

However, when the two signals are discretely generated and are transform-limited (are not chirped in time), the only way to broaden the signals

in time is to reduce their bandwidth. To achieve sufficient power-spectral-density to cause efficient conversion, the pulses must either generated by a low-bandwidth laser, or much power will be wasted in filtering a higher bandwidth signal. Unfortunately, the range resolution in OCT, when temporal ranging is used, is inversely proportional to the illumination bandwidth. This requirement for high bandwidth conflicts with the requirement for small bandwidth for resonance specificity. It would be desirable to come up with an alternate oscillator configuration that would preserve the resonance specificity of the low bandwidth pulses, but actually utilize high bandwidth signals.

Since the nonlinear excitation of the resonant molecule depends more on the beats produced than on the bandwidth used to produce them, it would be desirable to take a broadband pulse and reshape it into a signal with the required beat frequency. Recent advancements have made pulsed sources of very large bandwidth. Some of the methods to do this are high-bandwidth Ti-sapphire oscillators, dispersion compensated mirror Ti-sapphire oscillators, double chirped-mirror Ti-sapphire oscillators, and continuum generation sources. The optical field produced by these sources can be shaped into a field with the beats at the required frequency.

One such method that has been demonstrated in the literature is shown in Figure 4. A source of laser pulses from a laser source **401** is sent into a Fourier-plane pulse shaper **403** that utilizes a spatial-light-modulator (e.g. liquid crystal modulator or acousto-optic modulator). The Fourier-plane pulse shaper enables each frequency in the pulse to have its phase and/or amplitude altered. By applying the correct phase and amplitude to each incoming frequency, the incoming signal can be convolved with an essentially arbitrary signal. The pulse shaper is set up to reshape the incoming pulse by applying a period phase or amplitude perturbation in the Fourier domain with period $\Delta\omega$/N, where N is a positive integer. This will transform a single pulse into a train of pulses that are separated in time by $2\pi/\Delta\omega$. If only a phase perturbation is used, the power of the signal can be maximally preserved. The larger the integer N is, the longer the pulse train will be, and none of the bandwidth of the original pulse will be lost. However, most pulse shapers

have a limited number of controllable frequencies, limiting the practical size of N.

One advantage of spatial-light-modulator based pulse shapers is that there is typically a wide range of pulse shapes that can be achieved, and the spatial-light-modulator can often be controlled automatically by a computer. The computer can then adjust the spatial-light-modulator to achieve maximum signal from the sample in a feedback loop. This may allow automatic correction of dispersion or aberrations introduced by the optics of the system, and will permit more flexibility in probing the molecule because of the tunability of the pulse shapes.

The pulse-shaper in the "Pulse-shaper type NIVI oscillator" **400** is well described in the literature. It consists of two diffraction gratings, which disperse and recombine the frequencies, two lenses that focus each frequency to a point and recollimate each frequency, and a pulse shaper placed at the focal plane to separately operate on each frequency. The pulse shape is altered by dispersing each frequency to a separate angle, and then imaging each frequency to a separate point on the spatial-light-modulator. Alternatively, an etalon may be used to shape the amplitude of the pulse periodically. Unfortunately, while this would be simpler, it modifies the spectrum of the pulse and therefore introduces artifacts into the NIVI image.

An alternative method is to take a pulse and impart a linear chirp to it. A linear chirp turns a pulse into one where the frequency rises or falls at a linear rate as a function of time. This rate is characterized by a constant α, which is the change in frequency per unit time. It is called "chirped" because of the noise of the equivalent sound wave. If two copies of the chirped pulse are created, delayed with respect to each other by imparting a variable time delay **407**, and recombined, the resulting pulse will have two simultaneous frequencies that will rise or fall together at the same linear rate, but always be separated at a given instant by the same frequency. If this separation frequency is chosen to be $\Delta\omega$, then the envelope of the pulse will be modulated by beats of this frequency. This method is especially convenient because the probed resonance frequency can be adjusted easily by adjusting

the time delay between the two chirped pulses, which is relatively easy and inexpensive. This will enable a NIVI instrument that can be rapidly and easily adjusted to scan a wide range of molecular resonances. Systems based on tunable frequency sources will likely be much more difficult to dynamically change reliably and often.

The ability to linearly chirp a pulse is well known in the literature. It can be accomplished with a pulse shaper **405** having a combination of prisms, diffraction gratings, lenses, mirrors, and/or dispersive materials. Combinations may be required to ensure that the resulting chirp is linear and does not contain significant amounts of higher-order dispersion. Higher-order dispersion would limit the resolution to which the resonance could be addressed and exclude other nearby frequency resonances. In a typical Chirped CARS NIVI setup **402**, the chirp rate required would be fixed and the chirp rate should need little or no adjustment in the field. Measuring devices such as Frequency Resolved Optical Gating can test whether a chirped pulse is linearly chirped.

When using high-bandwidth excitation for CARS, it is important to filter out the entire bandwidth of excitation before detection so it does not interfere with detection of the emitted anti-Stokes light (Stokes for CSRS), because the nonlinear emission can not be easily distinguished from the much larger linearly scattered excitation light. However, this linearly scattered light contains the same structure that conventional OCT imaging does, and may be used to measure this information at the same time that a NIVI image is recorded. This will be convenient for superimposing OCT and NIVI data onto the same image, because the acquisition of both types of data can be designed into the same instrument. This can be implemented in practice by using a dichroic beamsplitter to separate the excitation and response radiation, and measuring the cross-correlation of the two frequency bands separately using a cross-correlation demodulator.

With high-bandwidth sources where the entire bandwidth need not be utilized to produce the excitation field, it is possible to use the upper end (for anti-Stokes) or lower end (for Stokes) of this bandwidth as a reference field,

eliminating the need to separately generate a reference field. However, the frequencies of the reference must occupy the same band as the received CARS/CSRS light from the sample. Some sources, especially continuum generation sources, will likely generate much more bandwidth than needed for pulse shaping and therefore will probably have this extra bandwidth available for this use. While a different process from CARS/CSRS typically generates this light, it will likely remain phase-coherent with the CARS/CSRS light and therefore should be useful as a reference. Phase-coherence depends on the mechanism of pulse/continuum generation and therefore its phase-coherence stability properties of a particular source type must be established before it is suitable for this purpose. A dichroic beamsplitter may be used to separate the frequency band corresponding to the response radiation from the oscillator energy, so that it may be utilized as a reference signal.

It is also possible to simultaneously stimulate the excitation of several resonances if the sample is illuminated with the appropriate pump and Stokes/anti-Stokes beams. For example, in CARS a narrowband pump signal and a wideband Stokes can be used to address many resonances simultaneously. This is called multiplex CARS and can be extended to CSRS with a broad anti-Stokes wavelength range. This may be used to measure the presence of several molecular resonances simultaneously in the sample. In addition, if several excitations can be produced in the same molecule simultaneously, the molecule will evolve to various quantum states depending on the relative amplitude and timing between the CARS/CSRS stimulating signals for each resonance. This may be produced by sending in multiple pairs of Stokes/anti-Stokes wavelengths and pump beams in with varying time delays between them. By varying the time delay between excitations, the molecule can be made to prefer Stokes or anti-Stokes emissions from a particular quantum state. This way, the amount of anti-Stokes radiation generated from each quantum state could be potentially probed to identify the molecule.

In general, the spatial-light-modulator system of Figure 4 could be used to produce more general pulse shapes than a series of beats at a single

resonance frequency. By using a more complicated pulse shape, several bonds present in a molecule can be coherently excited, and energy transferred between them in a coherent fashion. Because each molecule has some difference between the bonds presents and their relative orientation (and therefore the coupling in the quantum wave functions between them), pulses can be shaped that will selectively transfer energy between the states for a particular molecule, and not be selective for other similar molecules. In this way, the emission of stimulated Raman scattering or another coherent scattering process can be made more specific than just every molecule possessing a bond of a particular energy. With an automatically controlled pulse shaper, such as those based on spatial light modulators, feedback can be employed where the computer can test various pulse shapes, measure the resulting emitted light temporal signal using the demodulator, and progressive reshape the pulse to optimize the signal from the target molecule and exclude other molecules. Once a useful temporal field shape for stimulating a molecule has been found, it can be stored in a database and later used for identifying that molecule in the future.

2. **Reference Generator.**

The reference generator takes a portion of the signal produced by the oscillator and converts it to a reference signal. This reference signal is later used to demodulate the sample signal. The reference generator is a nonlinear process that converts light in the illumination bandwidth to light in the sample's emission bandwidth, so that interference can occur between them. This nonlinear process may or may not be CARS or CSRS.

A common implementation of the Reference Generator would be to focus the oscillator excitation into a sample of the same molecular species that one wishes to image. The reference signal should then be very similar to the same molecular species contained in the sample. This is because they are the same molecule, illuminated by nearly identical pulses, converting them to the output signal using resonant CARS or CSRS. The magnitude of the cross-correlation between these two signals should be great because of their

similarity. In addition, if there is variability of oscillator output, having the reference generator and sample contain the same substance will respond in similar ways, so that the cross-correlation signal can remain high despite fluctuations in the oscillator. The benefit of using the same molecule in the reference generator is that it is the molecule's signal that acts as its own "fingerprint" with which the cross-correlation can use to recognize the molecule in the sample. If more selective excitation processes than CARS are used, then using the same molecule in both reference and sample will help ensure that a reference signal can be generated for a given excitation field.

A nonresonant nonlinearity can be used as the reference generator as long as the peak power of the oscillator signal can excite a sufficient quantity of reference signal. High peak power can be maintained by not chirping the oscillator signal that is sent to the reference generator, while sending a relatively low peak power signal to the sample. Nonresonant CARS or CSRS can be used to generate an anti-Stokes or Stokes signal, respectively, from a medium that does not necessarily have a resonance at the frequency of the target molecule. The benefit of this is that the medium may not have to be changed each time a different molecular species is scanned for, because otherwise a medium with a resonance at that wavelength would need to be chosen. Also, this species can act as a standard signal source against which the return signals from many samples can be compared. The nonresonant CARS can be implemented by focusing the excitation radiation into a sample of liquid that produces a CARS/CSRS signal in the same frequency band as that generated from the sample. For example, benzene will generate a CARS anti-Stokes signal in the 3000-3100 cm^{-1} frequency band.

Continuum generation is another type of nonresonant nonlinear process that can be used in the reference generator. A sufficiently high peak power pulse is focused into a medium, where it excites a broad bandwidth of frequencies to be produced. If the produced frequencies overlap the emission frequency band produced in the sample, this portion of the continuum can act as a reference signal. The generated continuum must be created by a

mechanism that is sufficiently stable to not be overly sensitive to fluctuations in oscillator intensity. An unstable reference signal will result in noise in the cross-correlation signal. The benefit of continuum generation is that is likely to create a broad bandwidth that will have signal in the emission bandwidth of the sample, so that the continuum need only be filtered for the needed frequency band. Also, if the oscillator employs continuum generation, it may already generate light within the emission bandwidth that can be used as a reference generator, eliminating a separate nonlinear process in the reference generator step. Some examples of materials used for continuum generation materials are optical glass, fused silica, calcium fluoride, sapphire, ethylene glycol, water, high numerical aperture optical fibers, photonic crystal optical fibers, microstructured optical fibers, dispersion-shifted optical fibers, and gas cells (e.g. cells filled with helium, argon, or nitrogen).

Other candidate processes for nonresonant nonlinear reference generation include second and higher harmonic generation, stimulated Raman scattering, sum and difference frequency generation, optical parametric amplification, four-wave mixing, and self phase modulation.

Figure 5 shows an implementation of a reference signal generator **500**. The concentration optics **503** are typically implemented as some combination of lenses and mirrors. The concentration optics may also require some combination of frequency dispersive elements such as prisms, diffraction gratings, pulse shapers, and dispersive materials to prepare the temporal shape of the signal for nonlinear generation. Concentration in space and time may be necessary because the nonlinear processes are power sensitive, and depends on the strength of the nonlinear process. When the light **501** enters the nonlinear medium **505**, it undergoes conversion to a frequency band coinciding with the frequency band of the response signal from the sample. This nonlinear medium may be one of the media mentioned above, either a sample of a target molecule, a solvent, or a continuum generation medium. After exiting the nonlinear medium, the reference signal **509** is collected by reference collection relay optics **507**, where it is sent to the demodulator where it is combined with the sample signal. The collection relay optics are

usually implemented as some combination of lenses and mirrors that collimate the reference field radiation. This reference field should be characterized to find its temporal structure by instruments such as Frequency Resolved Optical Gating, cross-correlation with another known signal, nonlinear sonograms, or nonlinear autocorrelations/cross-correlations.

3. **Microscope.**

The microscope delivers the excitation radiation from the oscillator to the sample and collects the resulting sample emission. Inside the sample, a coherent nonlinear process such as CARS or CSRS takes place that, in the presence of a molecule of interest, will emit the sample field in response to the excitation field. The sample field is collected by the microscope and then propagated to the demodulator, where the sample field can be estimated from the measured cross-correlation between the sample field and reference field.

Microscope systems can be differentiated by various implementation choices. They can either illuminate one (serial scanning) or many points (full field imaging) at a time on the tissue. If the pump and Stokes (anti-Stokes) beams of the excitation field are separated in frequency, they can be sent in either separate (non-collinear) or identical (collinear) angles into the sample. The temporal delay of the response radiation relative to the reference may or may not be used to range molecular constituents in the tissue. The response radiation can be collected in the forward scattering (forward CARS/CSRS) or backward scattering (epi-CARS/CSRS) directions.

The microscope measures spatially resolved molecular density by illuminating various points on the tissue with the oscillator field, collecting the emitted sample field, and recombining with the reference field in the demodulator. One point in the tissue may be illuminated at a given time, resulting in serial or raster scanning of the molecular density through the tissue. Alternatively, a line or a complete plane of points may be illuminated, so that data may be acquired from many points in parallel. Illuminating and measuring the radiation from an entire plane of points is called full field imaging. At the time of this writing, full field imaging is seldom used because it requires an array intensity detector such as a charge-coupled-device (CCD)

to simultaneously measure the demodulated signals of all of the illuminated points. Unfortunately, as of this writing CCD arrays produce thermal dark noise at each pixel, and also have a relatively limited dynamic range of measurable intensity values. Demodulated interference signals often require very high dynamic range detection. It is conceivable that future CCD or other types of focal plane arrays (e.g. CMOS arrays) may overcome these limitations. Full field imaging also requires that the tissue be illuminated by larger amounts of power because measurable signal must be produced for an entire area rather than just one point. Since this is more likely to result in tissue damage, full field imaging will probably be used when speed of acquisition is paramount.

Figure 6 shows three examples of full field CARS configurations. The "beam delivery optics" are usually implemented as some combination of mirrors and lenses that deliver beams that illuminate a wide area or line on the sample. For all of these microscope configurations, beam delivery and collection optics will typically utilize a microscope objective. The "response field collection optics" **601** are similarly implemented as a combination of lenses and mirrors that relay the response field to the demodulator so that it may be recombined with the reference and detected. The noncollinear full field CARS **602** delivers the pump **603** and the Stokes **605** beams (assuming they are separate) at separate angles, so that the response is separated in angle from the illumination by an angle given by phase-matching conditions. The collinear geometry **607** sends the pump and Stokes radiation in the same directions (or in a single field if they can not be separated) and collects the radiation in the same direction, which can be discriminated with a dichroic beamsplitter. The epi-CARS geometry **611** collects the backscattered radiation, usually through the same objective optics that the sample is illuminated through. The epi-CARS can be discriminated from the illumination with a dichroic beamsplitter or an interference filter.

If collinear CARS is used, where the pump and Stokes (anti-Stokes) beams from the excitation beams overlap, then the response field **609** will overlap the excitation beams. Then the response field frequency band should

not completely overlap the excitation frequency band, so that a spectral filter may distinguish between the excitation and response fields. If non-collinear CARS is used, then the response beam can be sufficiently angularly separated (as determined by the phase-matching criterion) from the excitation radiation to be filtered by a spatial filter. However, in the case of non-phase-matched CARS, such as epi-CARS, the interaction CARS volume must be small enough so that the response is radiated effectively isotropically, so that spatial filtering is unavailable and spectral filtering should be used.

The other, more commonly used alternative is serial point scanning. Serial scanning tightly focuses the oscillation signal into the tissue to create a very small volume where peak power is maximized. The focusing is usually achieved using a microscope objective. This focus is then scanned through a 3-D set of points in the tissue, and the sample signal gathered from each point is demodulated to produce a 3-D NIVI image. Since nonlinear processes are power sensitive, efficient CARS/CSRS occurs only at the focus. If the sample is small enough, the focus may be scanned through the sample by translating the sample in all three dimensions. However, it is not feasible to move large samples such as human subjects this way. In this case, the beam focus can be moved in the transverse direction by steering the beam, perhaps using galvanometer rotated mirrors, acousto-optic modulators, or translating the lens assembly. The depth may be scanned by mechanically adjusting the distance between the lens and the tissue, perhaps using a lead-screw translator and/or a piezoelectric transducer. Since the signal can be excited at only one point at a time, one can be sure that the resulting measured sample signal is due to the interference of emissions of radiation produced in that volume only. This can be a benefit in NIVI when high phase resolution is required because one is assured that any measured phase shifts are not due to interference between molecules at disparate spatial locations. This higher phase resolution may be used to better differentiate between similar molecular species. Serial point scanning typically utilizes a single photodetector or a small number of photodetectors at the demodulator, which has the benefit that the dark current of a single photodetector is usually less

than that of an entire CCD array, and a single photodetector can also typically handle a higher dynamic range of measurements.

Figure 7 shows the geometry of translated serial-port scanning configurations. In all of these configurations, the sample is on a translator that moves the sample through the focus to form an image of the molecular density at various points. The translator could be a three-axis linear screw drive translator, or piezoelectric translator, or a combination of these. The beam delivery optics **701** focuses the pump **703** and Stokes (anti-Stokes) **705** beams onto the tissue at a point of interest, and the response field collection optics **707** gathers the generated anti-Stokes (Stokes) field **709** from the tissue. In the non-collinear geometry **700**, the pump and Stokes illuminate the point of interest at different angles, so that the anti-Stokes will emerge at a third angle given by phase-matching, so that the anti-Stokes is spatially separated from the illumination. In the collinear geometry **702**, the illumination and response fields emerge overlapped, so that they must be discriminated by frequency (e.g. using a dichroic beamsplitter). Finally, in the epi-CARS geometry **704**, the backscattering response fields are collected, often by the same optics through which the illumination is projected, and can be separated with a dichroic beamsplitter.

Alternatively, the focus can be moved and the sample can be left stationary. This can be accomplished by a combination of tilting the illumination beams before it enters the beam delivery optics and/or translating the beam delivery optics around (Figure 8). Translating the collection optics and/or tilting the exiting response beam with a beam steering optics **801** can capture the exiting response field. The beam steering optics changes the direction of the incoming beam **803**. This can be implemented, for example, by a galvanometer-scanned rotating mirror. By changing the direction of the beam before it enters the beam delivery optics **809**, the position of the focus in the sample can be changed. The translation for the beam delivery and collection optics **807** can be implemented with a piezoelectric and/or a linear screw-drive translation stage. Translating the beam delivery and/or collection optics moves the focus with the optics through the sample. These two

mechanisms can be combined to enable the three-dimensional translation of the focus through the sample. This configuration is especially convenient for the epi-CARS configuration, because the same beam steering optics and beam delivery optics **809** can be used to collect the response field **805**. For these configurations, a compensating delay may need to be incorporated into the demodulator because steering the beam and/or translating the delivery and collection optics change the travel time for the illumination and response signal through the microscope section.

All of the previous scanning modes, full field imaging, translated serial-point scanning, and beam-steering serial-port scanning, use the spatial location of the illumination beam to differentiate between response signal gathered from various points in the sample. This confinement method is the same as that used by multiphoton microscopy or CARS microscopy. Other technologies such as Optical Coherence Tomography and Optical Coherence Microscopy use temporal ranging in addition to spatial confinement to further isolate the contributions of signal from various points in the sample. Because of the heterodyne nature of NIVI, this scanning mode is also available. It may be attractive for *in vivo* imaging because it will enable scanning tissue without translating the microscope objective or sample, and may scan faster because there are mechanisms for scanning the temporal delay much faster than translating objective optics. The phase measurement capability of NIVI is useful for both temporal ranging and molecular species identification this way.

Figure 9 shows two configurations of NIVI **900** and **902** utilizing temporal ranging. Temporal ranging is achieved by measuring the interference of the reference and response signals for various relative temporal delays. The temporal gating configuration superficially resembles the other epi-CARS configurations. However, unlike previous scanning modes, the depth-of-field of the focusing of the illumination from the beam delivery optics in the tissue is set to be long, because the temporal gating will be used to discriminate between molecular constituents within the depth-of-field. This temporal ranging utilizes epi-CARS/epi-CSRS because the backscattered response signal is collected from the sample by the response

field connector so that the signal delay into the tissue can be timed. Because CARS/CSRS is not phase-matched for the backwards direction, the generation of a backscattered signal will only occur efficiently for small particles or edges of particles with a compatible resonance. The chief difference between this configuration and the other epi-CARS configurations is that the interference signal is scanned as a function of relative delay, and an interference signal maximum indicates the presence of a molecular species at a particular depth in the medium (corresponding to that time delay). In the full field configuration **900**, an entire plane of points is interfered with the reference signal to produce simultaneous measurements of the cross-correlation signal for the entire plane **909**. By scanning the delay, the molecular density of various planes can be measured. For the beam-steered temporal gating setup **902**, the beam is scanned through the sample by tilting the beam through the beam steering optics **903**. Beam steering provides lateral displacement of the beam, and temporal gating provides depth information so that a three-dimensional volume is scanned. The sample can be translated laterally also to scan the beam. The implementation differences between the temporal gating configurations are in the choice of depth-of-field of the illumination and collection optics, and in the demodulator. The demodulator must be designed to scan a sufficient temporal interval to capture the interference signal between the reference and response signals. This temporal interval is typically between 500 microns to 10 mm.

4. **Demodulator.**

The purpose of the demodulator section is to decode the response signal from the sample so that it may be measured by relatively slow electronic equipment (slow compared to the oscillations of the electric field of the response signal). The magnitude of this demodulation signal will be related to the density of a molecular species of interest in the tissue. With knowledge of the molecular density at each point, a molecular density map, or NIVI image, can be presented to the user. This demodulation is implemented as the cross-correlation of the response signal relative to a generated

reference signal. Outlined below are various optical configurations that produce this cross-correlation signal.

There are various design choices that are made when choosing a demodulator. First, one needs to know whether or not full field microscopy is used. Also, one must decide whether the cross-correlation will be measured one sample at a time, or many samples in a single instant.

Figure 10 shows an example of a cross-correlator that can be used to demodulate the full field CARS signal, measured for example as in Figure 6. The response field is collected and relayed by the "response field collection optics" of Figure 6 to the Response Field **1001** of Figure 10. The reference field **1003** is produced by the Reference Generator and relayed to the cross-correlator. The reference is delayed relative to the response field, and they are mutually overlapped using a beamsplitter **1005**. The "response field spectral filter" **1007** filters the recombined field for only the frequency band that contains the response field bandwidth, which removes any remaining oscillator signal. The filter can be implemented by a combination of interference filters, color glass filters, dichroic beamsplitters, or other frequency selective elements. The combined fields are imaged onto a focal plane array **1011**, such as a CCD, so that the CCD is the conjugate image plane of the sample plane. The imaging is achieved with the "relay imaging optics" **1009** which are some combination of lenses and mirrors. The intensity detected on the CCD corresponds to a single cross-correlation signal measurement collected from various points in the sample. The "variable attenuator" **1013** is adjusted to maximize the use of the dynamic range of intensity measurements of the CCD. As the focus in the sample changes (e.g. by translating the sample), and/or the adjustable delay **1015** is changed, the entire cross-correlation signal can be measured, forming a three-dimensional data set. The typical scan range length for high numerical aperture full field imaging will be up to 100 microns. The delay can be implemented as a mirror translated by a linear-screw-drive translation stage, or by piezoelectric actuation. This same cross-correlation can also be used for temporal ranging full field NIVI, as shown in Figure 9, with the only

difference being that the delay mechanism must be designed to scan a sufficiently long range of interest in the sample, typically from 500 microns to 10 mm.

The measurement of a cross-correlation signal for a serial-point scanning microscope is basically the same system as for full field imaging, except that a single photodetector can be used rather than an array photodetector. The adjustable delay, response field spectral filter, and relay imaging optics can all be implemented in similar ways to the full field case. The relay imaging optics will need only focus the combined response/reference signal onto the photodetector. In some cases, when only the magnitude of the cross-correlation signal at its peak needs to be measured, it may be desirable to dither the adjustable delay with a piezoelectric transducer a fraction of a wavelength, so that the magnitude of the cross-correlation signal peak can be demodulation with a multiplying mixer and low-pass filter (an electronic heterodyne demodulator). This configuration will be covered in more detail in part five. To measure the entire cross-correlation, the adjustable delay will be scanned over various time delays and the photodetector intensity signal measured.

In Figures 10 and 11, a variable attenuator **1013** is used to adjust the intensity of the signal that reaches the photodetector **1101** so that its dynamic range is not exceeded. Variable attenuators can be implemented with liquid crystal shutters, neutral density filter wheels, or rotating polarizers. The spatial filter **1103** is used on the response signal to filter out spatial inhomogeneities in the response signal that could reduce the depth of modulation at the photodetector. A spatial filter would typically consist of a telescope of two converging lenses, with a pinhole in the focal plane between the lenses to filter out power around the main diffraction focus.

When the oscillator produces pulses of a low repetition rate, so that for example high peak power can be employed, one may want to measure multiple samples of the cross-correlation signal simultaneously. This can be achieved by the configurations in Figure 12. These configurations have the advantage that, with a sufficient amount of pulse power and number of

photodetector array samples, the cross-correlation signal can be measured using a single pulse. Low repetition rate illumination can keep the peak power high while average power remains low.

The "linear photodetector array cross-correlator" **1201** expands the response and reference signals, and interferes them with an angle between the two beams. The beam expansion is achieved by, for example, a pair of converging achromatic lenses **1203** arranged in a telescope configuration. The points on the wavefront where the two signals combine will have various time delays between them. A cylindrical lens **1205** then focuses the beams into a line image on a linear photodiode array **1207**. This concentrates the signal to adjust for the narrow height of the detector pixels. Each intensity sample on the linear photodetector array indicates a sample of cross-correlation of the two fields with various relative time delays, with a constant intensity signal added. The recorder would then read the linear CCD signal so that a complete NIVI image can be assembled.

The primary disadvantage of this system is that the modulation of the intensity signal on the linear photodetector array is rather low, and the dynamic range of typical CCDs is likewise low, so that the signal cannot be measured to a high signal-to-noise ratio. To combat this, one can utilize the Fourier-transform cross-correlator **1209**. Rather than directly measuring samples of the cross-correlation, it interferes the two signals and measures their spectral decomposition. The beamsplitter **1005** combines the response and reference signals, which are filtered for the response bandwidth by utilizing a response field spectral filter **1007**. This signal is then filtered by a frequency dispersive element **1211** such as a diffraction grating that scatters each frequency to a different angle. The power of each frequency is then focused to a different pixel on the linear photodetector array **1215** by using a focusing element **1213** (typically a combination of lenses or mirrors). The samples of the intensity measured on the photodetector indicate the real part of the Fourier transform of the cross-correlation signal. The recorder may compute an inverse discrete Fourier transform of the linear photodetector array intensity samples to recover the cross-correlation. The Fourier

transform already performs the necessary Hilbert transform to infer the imaginary part of the cross-correlation signal. However, because it is an intensity measurement, the cross-correlation signal, a time-reversed version, and the autocorrelation of the response signal are superimposed in the intensity signal. By choosing the adjustable delay longer than the length of the cross-correlation temporal signal, the reconstruction of these signal components will not overlap in the time domain, and so the recorder device may distinguish the cross-correlation from its mirror image and the autocorrelation.

The adjustable delay may also be dithered a small amount to introduce a phase shift into each measured frequency component at the linear photodetector array. By utilizing at least three linearly independent phase shifts, the phase of each Fourier component can be known and the time domain cross-correlation computed using an inverse discrete Fourier transform. However, then the ability to measure the cross-correlation of a single pulse will be lost. Another possibility is to use a linear photodetector array with several rows (at least three) rather than just one row of pixels. By tilting one of the wavefronts slightly vertically with respect to the other, a small phase shift can be introduced in the measured cross-correlation signal between rows on the linear photodetector. With a sufficiently large phase shift, the complex amplitude of each Fourier component can be inferred from the intensity samples from each column on the linear photodetector by utilizing the discrete Fourier transform. Since all of the time delayed cross-correlation samples are measured simultaneously, this is a single-pulse measurement.

To minimize the effect of dark charge built up in the linear CCD detector, it is best to discard the charge as soon as possible before the pulse arrives, and read the CCD as soon as possible after the pulse arrives. The speed of readout should be as fast as possible given limitations in accuracy due to the readout noise.

5. **Record r.**

The recorder accumulates the samples of the cross-correlation signal gathered from various points in the sample, processes this data, and presents these as a human-interpretable image. It is usually implemented as a data acquisition digital computer with some method of automatic control of the adjustable mechanisms (such as delay lines, piezoelectric actuators, and galvanometer mirrors), analog-to-digital conversion, and some form of image output device such as a screen or printer. It may also have control of the oscillator itself, to automatically tune the wavelengths or bandwidths of output or control the rate or timing of pulse output. It may also control the delay lines or spatial light modulators in the pulse shaping mechanisms of Figure 4.

Typically, the recorder will scan the illumination through the sample and/or the adjustable time delay and measure the cross-correlation signal. Unless the Fourier-Transform cross-correlator is used, the intensity samples represent the cross-correlation signal with a constant level added that could be discarded. The magnitude of the cross-correlation signal indicates the presence of the target molecular species. For CARS/CSRS the magnitude of the cross-correlation signal is related to the second power of the molecular density of that species. By demodulating the magnitude of the cross-correlation signal, the molecular density can be estimated for the points on the cross-correlation signal for the areas from which the response signal was collected to form that cross-correlation signal. Both spatial confinement and temporal ranging can be used to differentiate between the signals due to molecular densities at different locations in the sample.

Figure 13 documents the relationship between the elements of the recorder **1300**. For serial-scanning or temporal-scanning configurations, the instructions from the operator are entered via a human interface device **1304** into the digital computer **1301**. The digital computer controls the scanned beam position in the sample using the galvanometer mirror angles of the galvanometer scanned mirrors **1303**, the oscillator **1302**, the adjustable delay line **1309**, the position of the microscope objective position **1305**, and/or the position of the sample translator **1307** to scan the illumination through the

sample. In general only a subset of these need be controlled to scan the three-dimensional volume of the tissue. If full field imaging is used, usually only the depth and/or the delay line need be scanned. The cross-correlation signal as measured by the intensity is converted to a voltage by the photodetector **1311**, which is in turn converted to a digital sample by the analog-to-digital converter **1313**. The digital computer varies the delay line and/or reads various pixels from the photodetector (if it is an array detector) to determine the cross-correlation signal. If a Fourier-transform cross-correlator is used, the computer will need to compute the inverse discrete Fourier transform of the signal to find its cross-correlation from the intensity samples. The computer then associates the magnitude of the cross-correlation signal with the molecular density, and assembles these magnitudes into a density map of the sample. This density map is stored in the storage device **1312** presented on the visual display **1315** and/or made into a physical representation with a printer or stereolithography device **1317**.

To aid in the measurement of the magnitude of the cross-correlation signal, the configuration **1400** of Figure 14 is suggested.

Because the digital computer **1301** can record only relatively slow signals, a dither oscillator **1401** may be used to add a small high-frequency dither signal into the adjustable delay line **1309** that produces a periodic perturbation of the delay in the signal of a magnitude usually of less than one wavelength. This same signal is multiplied by the received photodetector **1403** interference signal to demodulate it, and is low pass filtered through the low pass filters **1405** to remove harmonics of the dither frequency. Both the dither and its quadrature signal are demodulated, because these correspond to the real and imaginary parts of the complex cross-correlation amplitude. The analog-to-digital converter **1313** will then be directly measuring a quantity proportional to the complex magnitude of the cross-correlation for the delay line position. The computer can utilize the digitized real and imaginary cross-correlation components to display the amplitude and phase of the cross-correlation signal. The amplitude will correspond to the magnitude of the reflection, and the phase will correspond to the Doppler shift of the reflection.

If a very fast dither signal is desired (above 10 kHz), an electro-optic modulator or acoustical-optic modulator can be placed in the adjustable delay line system to rapidly modulate the delay a small amount.

The magnitude of the cross-correlation signal depends both on the density of molecules available to produce the response signal, and on the temporal structure of the signals radiated from the molecules. The temporal signal produced by the molecules can be inferred from the cross-correlation signal. If $f(t)$ is the temporal response signal produced by a molecule, and $g(t)$ is the known reference signal, the measured cross-correlation $\Gamma(\tau)$ is given by:

$$\Gamma(\tau) = \int_{-\infty}^{\infty} f(t)g(t-\tau)dt$$

If $\tilde{F}(\omega)$, $\tilde{G}(\omega)$, and $\tilde{\Gamma}(\omega)$ are the Fourier transforms of $f(t)$, $g(t)$, and $\Gamma(\tau)$ respectively, then $\tilde{\Gamma}(\omega) = \tilde{F}(\omega)\tilde{G}(-\omega)$. The function $\tilde{\Gamma}(\omega)$ can be computed from the Fourier transform of the cross-correlation and $\tilde{G}(\omega)$ can be computed from the measured reference signal. The Fourier transform $\tilde{F}(\omega)$ of temporal signal $f(t)$ can then be estimated by $\tilde{F}(\omega) = \tilde{\Gamma}(\omega)\tilde{G}(-\omega)^{*} / \left(\tilde{G}(-\omega)^{2} + N(\omega)^{2}\right)$ (change to equation), where $N(\omega)$ is an estimated power-spectral-density of the noise. An inverse Fourier transform of $\tilde{F}(\omega)$ yields $f(t)$.

The recorder can perform this computation and thus recover $f(t)$ for various molecular species. An unknown molecule may be identified by comparing its temporal signal to known molecular signals. In addition, by looking at the frequencies ω of phase $\arg\tilde{F}(\omega)$ where the phase changed rapidly, the resonance frequencies of the molecules of interest can be determined to high precision. A library of the temporal responses and resonance frequencies of various molecules for CARS excitation may be built up and used to identify unknown molecules *in vivo*. We also note that the same cross-correlation measurement configurations can be used to measure the OCT backscattered signal due to the scattering of the excitation radiation.

To do this, the excitation and response bands should be separated with a dichroic beamsplitter, and the cross-correlation setups implemented separately to eliminate photon noise from the excitation band being measured in the response band. Different cross-correlation setups may be used for each configuration, *e.g.* a single-pulse measurement system for the conventional excitation and a time delay cross-correlator as in Figure 11.

Embodiment

Randomness introduced into the measurement process may produce false positive indications of molecular densities, or obscure weak concentrations of molecular densities. This randomness has three sources: fluctuations in the oscillator, vibrations and air currents, and noise introduced by the photodetector.

Oscillator fluctuations are the hardest to characterize because the feedback mechanism of laser sources can produce large variations in the frequency, bandwidth, and output power of the pulses even for small perturbations of the oscillator. In addition, nonlinear processes such as self-phase-modulation, self-focusing (Kerr lensing), or continuum generation may also affect the pulse in ways that are sensitive to the power in the pulse. Therefore, it is desirable to keep fluctuations in pulse energy below a few percent, and keep frequency and bandwidth fluctuations under one percent.

Well designed state-of-the-art Ti-sapphire oscillators and regenerative amplifiers can meet these specifications if they are aligned and maintained within the guidelines set up by the manufacturer. It is important that the internals of the apparatus be shielded, as well as the paths of the beams, as much as possible from air currents and vibrations. This can be achieved by placing the apparatus on a platform with air flotation or shock-mounted legs. The oscillator apparatus should be enclosed in a rigid case with baffles that minimize the opening surface area to that required for the beams to enter and leave the case. It is best to use laser-diode or laser-diode-pumped solid-state sources to pump the oscillator to minimize pump fluctuations, but a well controlled ion laser is also usable.

A dispersion compensated mirror Ti-sapphire system can produce 200 nm or more bandwidth centered at 800 nm and is also suitable for the chirped-CARS or CARS utilizing a pulse shaper.

Another possibility is to employ a femtosecond diode-pumped erbium or ytterbium fiber laser. Fiber-based continuum generation sources of nanojoule energy have been demonstrated, and this source may be relatively inexpensive and compact. A pulse selector ("pulse picker") utilizing a Pockels cell or a fiber-based electro-optic modulator or electro-absorption modulator may be used to lower the average power while keeping the peak power high. A fiber amplifier can be used to increase the energy of the pulses. A fiber-based amplifier may provide a more portable, stable source.

If continuum generation is used, a source employing self-phase-modulation is preferable because of the short interaction length of the nonlinearity. For example, a microjoule energy pulse will broaden when focused into a fused silica, sapphire, calcium fluoride, or quartz medium. Self-focusing can be used to increase the peak power and minimize the interaction length. Self-phase-modulation in an optical fiber may be sufficient if it occurs over a short length of fiber. The fiber should be single mode in the bandwidth utilized to minimize nonlinear coupling of multiple modes. Higher peak power allows a stronger nonlinearity to be used and therefore minimizes the interaction length.

Photodetector noise is the other source of noise. There are two sources of photodetector noise: photon noise and thermal noise. Thermal noise may be eliminated by using a cooled detector, or by minimizing the integration time. If a line camera is used, the charge on the array can be expelled before the pulse is received and the signal can be read out immediately after the pulse is received. Also, the reference power should be balanced with the power received from the sample so that the dynamic range of the detector is utilized. Photon noise is fundamental to photodetection processes. Its effect can be minimized in several ways: using more illumination power to produce a larger CARS signal. Also, using the Fourier-

transform cross-correlator minimizes the photon noise effect by increasing the modulation when single pulse illumination is used.

Equipment Brand Names:

Coherent Laser (of Coherent, Inc.) 10W Verdi

pumps the

Kapteyn-Murnane Laboratories, LLC, titanium:sapphire seed (pump) laser

Which pumps the

Coherent Regenerative Amplifier (RegA 9050)

And uses the

Coherent Stretcher/Compressor.

A portion of the amplified light is sent to a

Coherent Optical Parametric Amplifier (OPA9450)

Which produces the second optical beam.

While we refer to a microscope (Leica) for beam delivery and holding the sample, any standard optics could be used. We are currently using a computer controlled 3-axis stage (Newport Corp.) to translate our samples under the imaging beams.

References:

1. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA, Fujimoto JG. Optical Coherence Tomography. Science 254:1178-1181, 1991.

2. Fujimoto JG, Brezinski ME, Tearney GJ, Boppart SA, Bouma BE, Hee MR, Southern JF, Swanson EA. Biomedical imaging and optical biopsy using optical coherence tomography. Nature Medicine 1:970-972, 1995.

3. Brezinski ME, Tearney GJ, Bouma BE, Izatt JA, Hee MR, Swanson EA, Southern JF, Fujimoto JG. Optical coherence tomography for optical biopsy: properties and demonstration of vascular pathology. Circulation 93:1206-1213, 1996.

4. Schmitt JM, Knuttel A, Bonner RF. Measurements of optical properties of biological tissues by low-coherence reflectometry. Appl. Opt. 32:6032-6042, 1993.

5. Sergeev AM, Gelikonov VM, Gelikonov GV, Feldchtein FI, Kuranov RV, Gladkova ND. *In vivo* endoscopic OCT imaging of precancer and cancer states of human mucosa. Opt. Express 1:432-440, 1997.

6. Tearney GJ, Brezinski ME, Bouma BE, Boppart SA, Pitris C, Southern JF, Fujimoto JG. *In vivo* endoscopic optical biopsy with optical coherence tomography. Science. 276:2037-2039, 1997.

7. Boppart SA, Bouma BE, Pitris C, Tearney GJ, Southern JF, Brezinski ME, Fujimoto JG. Intraoperative assessment of microsurgery with three-dimensional optical coherence tomography. Radiology. 208:81-86, 1998.

8. Hee MR, Izatt JA, Swanson EA, Huang D, Schuman JS, Lin CP, Puliafito CA, Fujimoto JG. Optical coherence tomography of the human retina. Arch. Ophthalmol. 113:325-332, 1995.

9. Puliafito CA, Hee MR, Lin CP, Reichel E, Schuman JS, Duker JS, Izatt JA, Swanson EA, Fujimoto JG. Imaging of macular disease with optical coherence tomography (OCT). Ophthalmology 102:217-229, 1995.

10. Puliafito CA, Hee MR, Schuman JS, Fujimoto JG. *Optical Coherence Tomography of Ocular Diseases*. Slack, Inc, Thorofare, NJ, 1995.

11. Schmitt JM, Knuttel A, Yadlowsky M, Eckhaus AA. Optical coherence tomography of a dense tissue: statistics of attenuation and backscattering. Phys. Med. Biol. 39:1705-1720, 1994.

12. Schmitt JM, Yadlowsky MJ, Bonner RF. Subsurface imaging of living skin with optical coherence microscopy. Dermatology 191:93-98, 1995.

13. Sergeev AM, Gelikonov VM, Gelikonov GV, Feldchtein FI, Kuranov RV, Gladkova ND, Shakhova NM, Snopova LB, Shakov AV, Kuznetzova IA, Denisenko AN, Pochinko VV, Chumakov YP, Streltzova OS. *In vivo* endoscopic OCT imaging of precancer and cancer states of human mucosa. Opt Express 1:432-440, 1997.

14. Profio AE, Doiron DR. Transport of light in tissue in photodynamic therapy of cancer. Photochem. Photobiol. 46:591-599, 1987.

15. Tearney GJ, Brezinski ME, Boppart SA, Bouma BE, Weissman N, Southern JF, Swanson EA, Fujimoto JG. Catheter-based optical imaging of a human coronary artery. Circulation 94:3013, 1996.

16. Tearney GJ, Brezinski ME, Southern JF, Bouma BE, Boppart SA, Fujimoto JG. Optical biopsy in human gastrointestinal tissue using optical coherence tomography. Amer. J. Gastroenterol. 92:1800-1804, 1997.

17. Tearney GJ, Brezinski ME, Southern JF, Bouma BE, Boppart SA, Fujimoto JG. Optical biopsy in human urologic tissue using optical coherence tomography. J. Urol. 157:1915-1919, 1997.

18. Boppart SA, Brezinski ME, Pitris C, Fujimoto JG. Optical Coherence Tomography for Neurosurgical Imaging of Human Intracortical Melanoma. Neurosurgery 43:834-841, 1998.

19. Bouma BE, Tearney GJ, Boppart SA, Hee MR, Brezinski ME, Fujimoto JG. High resolution optical coherence tomographic imaging using a modelocked $Ti:Al_2O_3$ laser. Opt. Lett. 20:1486-1488, 1995.

20. Drexler W, Morgner U, Kartner FX, Pitris C, Boppart SA, Li X, Ippen EP, Fujimoto JG. *In vivo* ultrahigh resolution optical coherence tomography. Opt. Lett. 24:1221-1223, 1999.

21. Tearney GJ, Bouma BE, Boppart SA, Golubovic B, Swanson EA, Fujimoto JG. Rapid acquisition of *in vivo* biological images using optical coherence tomography. Opt. Lett. 21:1408-1410, 1996.

22. Chen Z, Milner TE, Srinivas S, Wang X. Noninvasive imaging of *in vivo* blood flow velocity using optical Doppler tomography. Opt. Lett. 22:1119-1121, 1997.

23. Yazdanfar S, Kulkarni MD, Izatt JA. High resolution imaging of *in vivo* cardiac dynamics using color Doppler optical coherence tomography. Opt. Express. 1:424-431, 1997.

24. de Boer JF, Milner TE, van Germert MJC, Nelson SJ. Two-dimensional birefringence imaging in biological tissue by polarization sensitive optical coherence tomography. Opt. Lett. 22:934-936, 1997.

25. Tearney GJ, Boppart SA, Bouma BE, Brezinski ME, Weissman NJ, Southern JF, Fujimoto JG. Scanning single-mode fiber optic catheter-endoscope for optical coherence tomography. Opt. Lett. 21:1-3, 1996.

26. Boppart SA, Bouma BE, Pitris C, Tearney GJ, Fujimoto JG. Forward-imaging instruments for optical coherence tomography. Opt. Lett. 22:1618-1620, 1997.

27. Tearney GJ, Brezinski ME, Bouma BE, Boppart SA, Pitris C, Southern JF, Fujimoto JG. *In vivo* endoscopic optical biopsy with optical coherence tomography. Science 276:2037-2039, 1997.

28. Bouma BE, Tearney GJ, Compton CC, Nishioka NS. High-resolution imaging of the human esophagus and stomach *in vivo* using optical coherence tomography. Gastrointest. Endosc. 51:467-474, 2000.

29. Sivak MV Jr, Kobayashi K, Izatt JA, Rollins AM, Ung-Runyawee R, Chak A, Wong RC, Isenberg GA, Willis J. High-resolution endoscopic imaging of the gastrointestinal tract using optical coherence tomography. Gastrointest. Endosc. 51:474-479, 2000.

30. Li X, Boppart SA, Van Dam J, Mashimo H, Mutinga M, Drexler W, Klein M, Pitris C, Krinsky ML, Brezinski ME, Fujimoto JG. Optical coherence tomography: advanced technology for the endoscopic imaging of Barrett's esophagus. Endoscopy 32:921-930, 2000.

31. Boppart SA, Brezinski ME, Bouma BE, Tearney GJ, Fujimoto JG. Investigation of developing embryonic morphology using optical coherence tomography. Dev. Biol. 177:54-63, 1996.

32. Boppart SA, Brezinski ME, Tearney GJ, Bouma BE, Fujimoto JG. Imaging developing neural morphology using optical coherence tomography. J. Neurosci. Meth. 2112:65-72, 1996.

33. Boppart SA, Tearney GJ, Bouma BE, Southern JF, Brezinski ME, Fujimoto JG. Noninvasive assessment of the developing *Xenopus* cardiovascular system using optical coherence tomography. Proc. Natl. Acad. Sci. USA 94:4256-4261, 1997.

34. Boppart SA, Bouma BE, Pitris C, Southern JF, Brezinski ME, Fujimoto JG. *In vivo* cellular optical coherence tomography imaging. Nature Med. 4:861-864, 1998.

35. Pitris C, Goodman AK, Boppart SA, Libus JJ, Fujimoto JG, Brezinski ME. High resolution imaging of gynecological neoplasms using optical coherence tomography. Obstet. Gynecol. 93:135-139, 1999.

36. Pitris C, Jesser C, Boppart SA, Stamper D, Brezinski ME, Fujimoto JG. Feasibility of optical coherence tomography for high resolution imaging of human gastrointestinal tract malignancies. J. Gastroenterol. 35:87-92, 1999.

37. Boppart SA. *Surgical Diagnostics, Guidance, and Intervention using Optical Coherence Tomography.* Ph.D. Thesis. Harvard-MIT Division of Health Sciences and Technology, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1998.